

# Content-Adaptive Image Compressed Sensing Using Deep Learning

Liqun Zhong, Shuai Wan, Leyi Xie, Shun Zhang

School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China.

E-mail: zhongliqun0314@mail.nwpu.edu.cn, swan@nwpu.edu.cn

**Abstract**—This paper proposes a framework of content-adaptive image compressed sensing using deep learning, which analyzes the image content and adaptively allocates samples for different image patches accordingly. Experimental results demonstrate that the proposed framework outperforms the state-of-the-arts both in subjective and objective quality, especially at low sampling rates. For example, when the sampling rate is 0.1, 1-6 dB improvement in peak signal to noise ratio (PSNR) is observed. Moreover, the proposed work reconstructs images with more details and less image blocking effects, leading to apparent visual improvement.

## I. INTRODUCTION

The theory of compressed sensing (CS) has achieved remarkable progress [1], [2], [3], while its application in image compression is rather challenging. Regarding images, block-based CS (BCS) approaches are more suitable, as proposed in [4] which divides an image into small patches. Various improvements have been made, like [5] which combines the BCS with smoothed projected landweber reconstruction (BCS-SPL), and a multiscale variant of the original BCS-SPL (MS-BCS-SPL) reconstruction provided in [6]. In [7], multi-hypothesis prediction (MH-BCS-SPL) is employed to obtain the residuals of CS in the random projection domain. Moreover, [8] conducts compressed sensing by learning a gaussian mixture model from measurements.

Another direction upon CS is based on deep neural networks as described in [9], [10], [11]. The advantage of such an approach is that the measurement matrix and the nonlinear reconstruction operators can be jointly optimized during training, which therefore outperforms other available CS algorithms regarding image CS. Generally speaking, the performance of state-of-the-art is still limited, especially at very low sampling rates. For example, obvious blocking artifacts and loss of image details are always observed at very low sampling rates, seriously affecting visual perception.

In block-based compressed sensing, each image patch is generally assigned with the same number of samples. However, the sensitivity of human eyes to image distortion is different for different content. Inspired by this spirit, this paper proposes a way of content-adaptive image compressed sensing using deep learning, which can adaptively allocate appropriate number of samples for different image patches.

The rest of this paper is organized as follows. Section 2 provides a brief introduction on related work. The proposed framework and method are described in section 3. Experimen-

tal results are given in section 4. This paper concludes with a summary.

## II. RELATED WORK

### A. Block-based Compressed Sensing

In our work, the block-based CS sampling (BCS) will be used. In BCS, the image is decomposed into disjoint patches of size  $B \times B$ . Each patch is sampled separately. Suppose that  $\mathbf{x}_j \in \mathbf{R}^{B^2}$  is a vector representation of patch  $j$  of input image, using a raster-scan fashion. Since  $\mathbf{x}_j$  can be expressed as  $\mathbf{x}_j = \Psi\theta_j$ , in which  $\theta_j$  is nearly sparse, the corresponding measurement  $\mathbf{y}_j$  can be expressed as:

$$\mathbf{y}_j = \Phi_B \mathbf{x}_j = \Phi_B \Psi \theta_j = \mathbf{A} \theta_j. \quad (1)$$

Here,  $\Phi_B$  is an  $M_B \times B^2$  orthonormal measurement matrix, in which the sampling rate  $R = M_B/B^2$ .  $\Psi$  is a certain transformation matrix, and  $\mathbf{A}$  is a sensing matrix.

Compressed sensing theory indicates that in order to reconstruct  $\mathbf{x}_j$ , the sensing matrix  $\mathbf{A}$  needs to satisfy restricted isometry property (RIP) [12]. The signal sparsity can be described by  $\ell_0$  norm, whereas it turns out to be an NP hard problem. Regarding the situation that  $\ell_1$  norm is equivalent to the  $\ell_0$  norm under certain conditions [2], the  $\ell_1$  norm can usually be used instead. Mathematically, CS reconstruction means to solve the following optimization problems:

$$\min \|\theta_j\|_1 \text{ s.t. } \mathbf{A}\theta_j = \mathbf{y}_j. \quad (2)$$

### B. Compressed Sensing Based on Neural Network

References [9], [10], [11] have presented some deep neural networks-based solutions to the problem of CS image sampling and reconstruction, in which fully-connected or convolution network (CNN) are utilized to conduct CS sampling and reconstruction. The general CS structure based on network can be illustrated in Fig. 1:

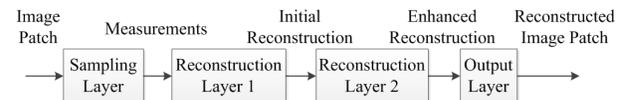


Fig. 1. Compressed sensing framework based on deep neural networks.

The first layer in Fig. 1, as a sampling layer, projects the input image patch of size  $B \times B$  into a vector of dimension  $RB^2$ , which is also the measurements obtained. Here  $R$

represents sampling rate. The second and third layer followed by an activation function RELU [13] are mainly used to conduct reconstruction operation. More in detail, the second layer usually recovers some structure information from CS measurements. For example, in [9], [11] a fully-connected layer is employed to recover such information. Then the third layer extracts more image details to enhance image quality. Fully-connected layers and CNN layers are both available ways in this part. The last layer, as an output layer, outputs the reconstructed image patch of size  $B \times B$ . Especially, reference [9] provides a network structure which contains a total of four fully connected layers, where fully connected networks perform block-based linear sensing and nonlinear reconstruction. Considering its excellent performance, we employ this framework in our simulation experiment latter.

### III. PROPOSED WORK

Although deep neural networks have greatly improved CS in sampling and reconstruction, the related subjective and objective quality is still not satisfactory when the sampling rate is relatively low. In non-adaptive CS sampling, the number of sampled CS-measurements for each patch is always the same at a fixed sampling rate, without considering the various contents of different patches. As is known, a natural image generally contains different contents in different regions. In this section, we focus on the allocation of samples for various patches, with reference to their contents, thus achieving a higher sampling efficiency in the quality of images. Our framework is illustrated as in Fig. 2.



Fig. 2. The proposed framework.

#### A. Content Analysis

In order to investigate the relationship between sampling rates and image quality in CS based on deep neural networks. Given an image, we utilize block-based compressed sensing, in which the image is divided into image patches of size  $B \times B$ . In our method, the mean square error (MSE) decline curve is analyzed to distinguish the smooth and complex image patches. In general, the slower the MSE declines, the smoother the corresponding image patch is, where human eyes are generally less sensitive to. Accordingly, few samples will be assigned to smooth areas and, correspondingly, the areas with more complex texture will be allocated with more samples, which are in line with human visual characteristics.

#### B. Sample Rate Allocation

On the basis of content analysis, we propose a content-adaptive sample rate allocation algorithm (CASRA) for network-based compressed sensing. It can be briefly described as: given a total number of samples assigned to an image, when an image is divided into patches, these patches can

be adaptively assigned with appropriate number of samples according to their characteristics. Provided with the assigned number of samples, the corresponding network models are employed for sampling and reconstruction. Fig. 3 illustrates the flow graph of the proposed method.

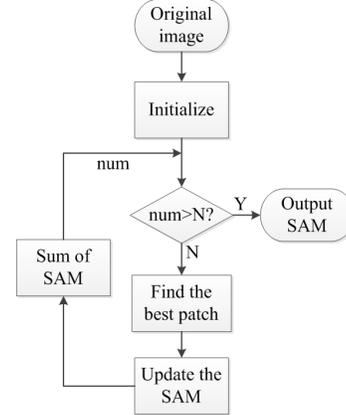


Fig. 3. CASRA algorithm flow chart, in which SAM means sampling map,  $N$  means the total number of samples.

How to find the best patch? Here we define the priority of an image patch as follows:

$$\Delta MSE_{h,w} = \max((MSE_{h,w}^{fp} - MSE_{h,w}^{fp+5})/5, (MSE_{h,w}^{fp} - MSE_{h,w}^{fp+10})/10, (MSE_{h,w}^{fp} - MSE_{h,w}^{255})/(255 - fp)), \quad (3)$$

where  $h$  and  $w$  represent the position index of an image patch,  $fp$  represents the number of samples to which the image patch has been allocated, and  $MSE_{h,w}^{fp}$  represents the MSE of the current image patch when the number of samples is  $fp$ . The larger the  $MSE_{h,w}$  is, the greater the corresponding priority is.

Given the same number of total samples, the proposed sample allocation method achieves a more rational sample allocation, resulting in improved image quality. In addition, since the CASRA guarantees the maximum amount of MSE descent as soon as possible when allocating samples, the final PSNR of restored images can be advanced significantly. Considering the number of samples of each patch is different, after samples of each patch are encoded, a stop symbol will be inserted to separate samples of different patches.

Specific steps are as follows:

- (1) Initialize by assigning 5 samples to each image patch uniformly to get an initial sampling allocation map (SAM).
- (2) Find the best patch with the highest priority and add 5 samples to it. Then update the SAM.
- (3) If the sum of the SAM is smaller than the total number of samples  $N$ , repeat step (2) to continue assigning samples. Otherwise, go to step (4).
- (4) The allocation is completed, the current SAM is the final allocation.

TABLE I  
PSNR FOR RECONSTRUCTED IMAGES, IN WHICH RATE IS SAMPLING RATE (DISPLAYED AS PSNR (dB)).

Rate	Method	Lena	Barbara	Peppers	Mandrill	Goldhill	Cameraman	Boat	Bridge	Mean
0.05	BCS-SPL-DDWT	24.62	21.42	24.68	19.67	24.98	22.81	22.77	21.1	22.76
	Gaussian ReconNet	25.93	21.3	25.85	19.07	25.11	24.81	23.18	21.09	23.29
	Adaptive ReconNet	29.14	23.37	29.59	20.98	27.92	28.37	26.07	23.59	26.13
	Network in [9]	29.74	23.77	30.79	21.21	28.72	29.4	26.83	23.94	26.80
	Proposed	<b>32.33</b>	<b>24.72</b>	<b>33.08</b>	<b>21.49</b>	<b>29.52</b>	<b>33.39</b>	<b>28.26</b>	<b>24.12</b>	<b>28.36</b>
0.1	BCS-SPL-DDWT	27.49	22.61	28.39	20.53	26.76	25.48	24.97	22.47	24.84
	Gaussian ReconNet	28.05	22.94	28.2	19.68	26.67	27.64	25.03	22.44	25.08
	Adaptive ReconNet	31	23.73	31.34	21.68	29.34	31.35	27.74	24.9	27.64
	Network in [9]	31.63	24.24	33.49	22.04	30.28	32.11	28.68	25.17	28.46
	Proposed	<b>36.09</b>	<b>27.46</b>	<b>36.26</b>	<b>23.14</b>	<b>31.96</b>	<b>38.45</b>	<b>31.36</b>	<b>26.1</b>	<b>31.35</b>
0.2	BCS-SPL-DDWT	31.08	23.81	32.55	21.75	29.02	29.45	27.9	24.14	27.46
	Gaussian ReconNet	31.13	22.88	31.43	20.86	28.82	31.41	27.79	24.3	27.33
	Adaptive ReconNet	34.11	24.38	34.68	23.08	31.56	35.17	30.5	26.78	30.03
	Network in [9]	34.97	25.11	36.65	23.89	32.93	36.8	31.79	27.36	31.19
	Proposed	<b>41.15</b>	<b>32.51</b>	<b>40.31</b>	<b>25.6</b>	<b>35.39</b>	<b>45.43</b>	<b>35.33</b>	<b>28.72</b>	<b>35.56</b>

Fig. 4 shows the sampling allocation for Lena, which confirms our suppose that smooth patches are generally assigned with less samples. The reason is that these patches always "fail" in competition due to their slow MSE decline so that fewer samples are assigned.



Fig. 4. Lena and its SAM map. Left to right: Original image; SAM map.

### C. Network-based CS

Besides sample rate allocation, considering the fact that network-based CS outperforms the state-of-the-art, we decide to utilize this method to conduct sampling and reconstruction. We reference the network model in [9], with four fully-connected layers, in which the first fully-connected layer is used as sampling layer, the remaining three layers are reconstruction layers. In our simulation experiment, we train 51 networks, corresponding to 5, 10, 15, ..., 255 samples respectively. We randomly select 20,000 images from the LabelMe dataset [14] and generate 5,000,000 patches of size  $16 \times 16$  as the training set, which is different from the test set used in experimental results.

## IV. EXPERIMENTAL RESULTS

This section provides the testing results of eight  $512 \times 512$  images, i.e. Lena, Barbara, Peppers, Mandrill, House, Bridge,

et. al. Images are divided into non-overlapping patches of size  $16 \times 16$ , and then independently sampled and reconstructed by BCS-SPL-DDWT [5], Gaussian ReconNet [10] (in which, CS measurements are obtained through a Gaussian matrix, then a fully-connected layer along with 6 CNN layers followed to recover images), Adaptive ReconNet [11] (in which, a fully-connected layer is used to sample CS measurements, then reconstruction layers are similar to the Gaussian ReconNet), network-based CS in [9] and our content-adaptive sample rate allocation (CASRA) respectively. Table 1 tabulates the PSNR and results in the range of R=0.05, 0.1, 0.2.

From Table 1, the network-based CS [9] is better than BCS-SPL-DDWT, Gaussian ReconNet and Adaptive ReconNet algorithms, which is proven to be an efficient way to conduct CS. It is also seen that compared with the existing compressed sensing methods, our proposed CASRA has obvious performance advantages. For example, when R=0.2, our proposed method gains an average of 4.3 dB and up to 7.4 dB in PSNR over the method in [9].

Fig. 5 illustrates Lena's subjective visual images and details. It is seen that the proposed method restores more image details so that the overall image quality has been improved. Therefore, we can conclude that the proposed framework is with advanced performance in both subjective quality and objective quality of the reconstructed CS images. More experiment results will be exhibited on the last page.

## V. CONCLUSIONS

In this paper, we propose a CS framework and a new method to adaptively allocate sampling numbers for different image patches, considering their various contents. The proposed work can automatically allocate fewer samples on smooth patches without affecting visual effects, while more samples are allocated to the others with more complex texture. As the result, the objective quality of restored images is improved, while the

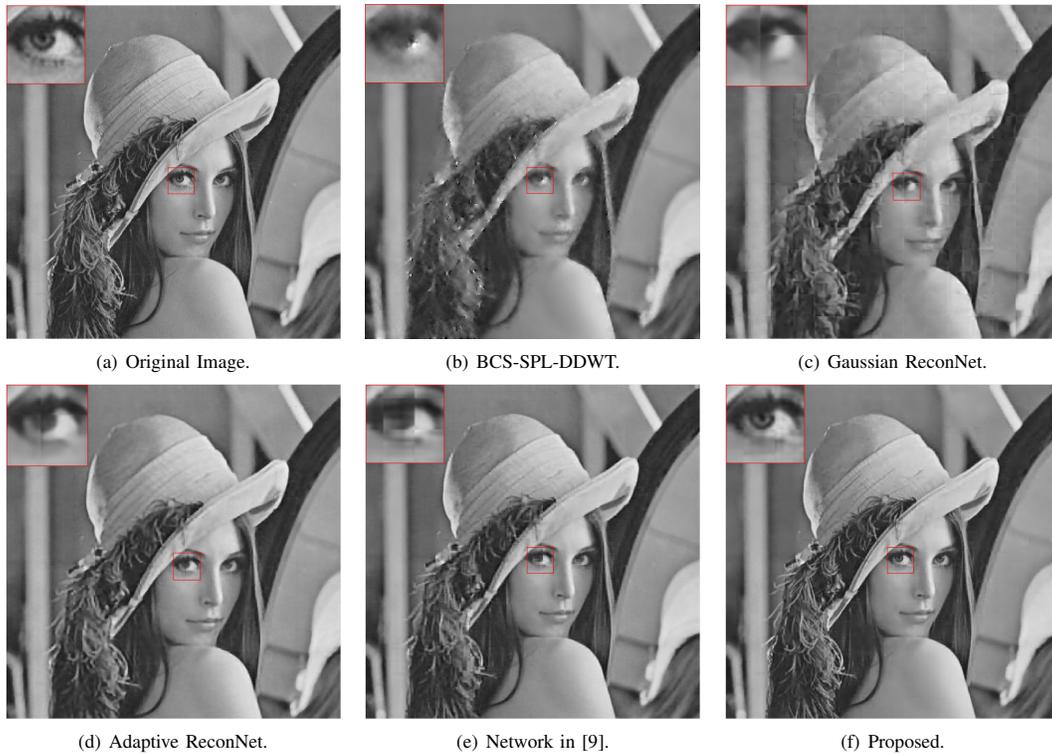


Fig. 5. Reconstruction of 'Lena' at sampling rate  $R=0.1$ , (a)Original image;(b)BCS-SPL-DDWT(27.49dB);(c)Gaussian ReconNet(28.05dB); (d)Adaptive ReconNet(31dB);(e)Network in [9] (31.63dB);(f)Proposed(36.09dB).

subjective quality is also enhanced since the proposed work complies with the characteristics of the human visual system.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation Research Program of China (Grant 61371089) and the Fundamental Research Funds for the Central Universities (Grant 3102016zy019).

REFERENCES

- [1] J. Romberg and T. Tao, "Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] E. J. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics*, vol. 7, no. 4, pp. 529–531, 2007.
- [4] J. E. Fowler, S. Mun, and E. W. Tramel, *Block-Based Compressed Sensing of Images and Video*, Now Publishers Inc., 2012.
- [5] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *International Conference on Image Processing*, 2010, pp. 2985–2988.
- [6] J. E. Fowler, S. Mun, and E. W. Tramel, "Multiscale block compressed sensing with smoothed projected landweber reconstruction," in *European Signal Processing Conference*, 2015, pp. 564–568.
- [7] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-sensing recovery of images and video using multihypothesis predictions," in *Signals, Systems and Computers*, 2011, pp. 1193–1198.
- [8] J. Yang, X. Liao, and et al., "Compressive sensing by learning a gaussian mixture model from measurements," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 106, 2015.
- [9] A. Adler, D. Boubilil, and Zibulevsky M., "Block-based compressed sensing of images via deep learning," in *International Workshop on Multimedia Signal Processing (MMSp)*, 2017, pp. 1–6.
- [10] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 449–458, 2016.
- [11] Xuemei Xie, Yuxiang Wang, Guangming Shi, Chenye Wang, Jiang Du, and Xiao Han, "Adaptive measurement network for cs image reconstruction," in *CCF Chinese Conference on Computer Vision*, 2017, pp. 407–417.
- [12] Candes and J. Emmanuel, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus - Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.
- [13] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*, 2010, pp. 807–814.
- [14] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: A database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157–173, 2008.

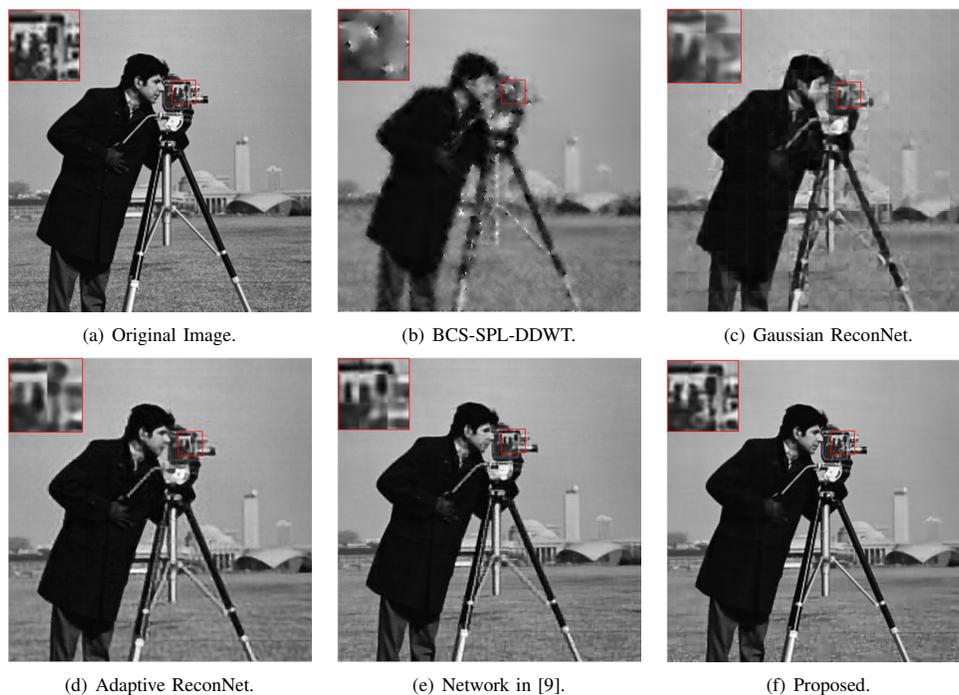


Fig. 6. Reconstruction of 'Cameraman' at sampling rate  $R=0.05$ , (a)Original image;(b)BCS-SPL-DDWT(22.81dB);(c)Gaussian ReconNet(24.81dB); (d)Adaptive ReconNet(28.37dB);(e)Network in [9] (29.4dB);(f)Proposed(33.39dB).



Fig. 7. Reconstruction of 'Barbara' at sampling rate  $R=0.2$ , (a)Original image;(b)BCS-SPL-DDWT(23.81dB);(c)Gaussian ReconNet(22.88dB); (d)Adaptive ReconNet(24.38dB);(e)Network in [9] (25.11dB);(f)Proposed(32.51dB).